METHOD FOR FABRICATING AN OLFACTORY RECEPTOR-BASED BIOSENSOR

This application is a continuation-in-part of application 09/057,181 filed April 8, 1998, the entire disclosure of which is incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to biosensors and, more specifically, to biosensors which have biomolecules attached to a thin film transducer.

BACKGROUND OF THE INVENTION

Chemoreception is an ancient sense system that enables organisms to detect chemicals in its environment. In humans, odor receptor cells are located in the nose. The biochemical receptors for the odorants are transmembrane proteins found in the membrane of receptor cell cilia. Olfactory receptor proteins (ORP) generally have seven non-intersecting helices. It is believed that conserved residues determine the orientation of each helix relative to the other helices. When the odor molecule binds to the receptor (in the transmembrane regions), it is believed that the receptor molecule changes shape. This apparently activates a G-protein on the intracellular surface of the cilia which in turn binds to a G-protein receptor on the ORP. (Olfactory G-protein receptors are one of the largest groups of G-protein coupled receptors described to date.) Olfactory G-protein linked receptors then trigger the biochemical synthesis of neurotransmitters which open cation channels that ultimately lead to action potentials and signaling, i.e. the sense of smell. In other words, the chemical stimulus is transduced into a neural event. The major path of olfactory transduction is shown in Figure 1 of the drawings.

There is currently a need for sensors which can detect ligands of the type which bind to olfactory receptor proteins. The goal, then, is to assign functional odorants to specific olfactory receptors and to develop useful sensors for detecting the presence of the odorants. It has been difficult in the past, however, to rapidly determine the secondary and tertiary molecular structures of ORP's having olfactory receptor binding domains specific

to selected ligands of interest. This is due in part to the complexity of ORP molecules. As will be understood by those skilled in the art, in an empirical analysis, a determination of putative binding domains is an extremely labor-intensive endeavor. It begins with identification and molecular cloning of genes that code for the receptor protein of interest. These genes are then expressed and the target protein is isolated and purified. Physical studies such as X-ray diffraction, neutron diffraction and electron microscopy are conducted to determine 2-D maps and 3-D structure; site directed mutagenesis is conducted to determine the position of residues for ligand binding. It would be desirable to provide a method which eliminates as many of these steps as possible.

Thus, it is an object of the present invention to provide a method for rapidly determining ORP candidates for use as receptors for preselected odorant molecules.

It is a further object of the present invention to provide a method for fabricating a biosensor which includes a layer of polypeptides that selectively binds a preselected odorant molecule.

SUMMARY OF THE INVENTION

In one aspect the present invention provides a method for making a biosensor capable of detecting a molecule, wherein the molecule is a ligand for an olfactory receptor protein. The method includes the steps of determining the amino acid sequence of a preselected olfactory receptor protein the secondary and tertiary structures of which are not known. Typically this step will be carried out by choosing an ORP from a database of ORP's which have been sequenced. In the next step the amino acid sequence of the ORP selected in the first step is compared to the sequence of G-coupled protein receptors having known secondary and tertiary structures. This step will typically be carried out by accessing a database of G-protein receptors having known primary, secondary and tertiary structures. Next, based on primary sequence homology, one or more G-coupled protein receptors is chosen as a candidate on which to predict the secondary and tertiary structures of the unknown ORP. In the next step, the secondary and tertiary structures of the unknown ORP are approximated based on the known structures of the G-protein receptor

selected through sequence homology comparison in the prior steps. The approximated secondary and tertiary structures of the unknown ORP are then analyzed using conventional modeling techniques to identify likely binding domains for the ligand of interest. A polypeptide is then synthesized having the primary sequence of the most likely binding domain for the ligand. These polypeptides are attached to a transducer. The resultant biosensor is then tested by exposing it to the target ligand and determining binding efficiencies. By identifying and testing a number of polypeptides in this manner, high affinity biosensors can be rapidly fabricated.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a diagram illustrating the major pathway of olfactory transduction.

Figure 2 is a flow chart illustrating the modeling steps of the present invention.

Figure 3 is a perspective view of a transducer made in accordance with the present invention.

Figure 4 is an amino acid sequence for ORP P30955.

Figure 5 is a table illustrating frequency changes resulting from attachment of ligands to a polypeptide made in accordance with the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring now to Figure 2 of the drawings, an olfactory receptor protein which has been sequenced is selected. Of course, it may be desirable in some cases to actually clone, express, isolate and sequence a new ORP; however, in most instances an ORP will be chosen from a sequence database having the primary amino acid sequence of various ORPs. One preferred database for use in the present invention is available on the ExPASy server of the Swiss Institute of Bioinformatics. Other similar databases or print sources may be equally suitable.

Once the ExPASy server has been accessed, the file entitled "SWISS PROT and TrEMBL- protein sequences" is opened. The ExPASy server is open to the public and may be accessed via the Internet. Next, using the keyword search feature of this file, the key words "olfactory receptor" may be used to create a subset of sequences of olfactory receptor proteins. An ORP is then selected, the sequence of which is to be used in the practice of the invention. The known sequence is displayed along with additional information on the ORP such as EMBL cross references, length and molecular weight. The amino acid sequence information is generally subdivided into potential extracellular and cytoplasmic domains.

In the next step of the invention the sequence of the ORP of unknown secondary and tertiary structures is compared to sequences of proteins having known sequences and known secondary structures. Most preferably, the database of proteins with known secondary structures is comprised of G-coupled receptor proteins. It will be appreciated by those skilled in the art that olfactory receptor proteins are a class of G-coupled receptor proteins. This comparison is preferably carried out using a publicly available database. Most preferably, the predicted secondary structure of the ORP under investigation is determined using the "PredictProtein" server of the "BIOcomputing 3D Modeling Unit Service" webpage (PredictProtein:B Rost (1996) Methods in Enzymology, 266:525-539; Url: http://dodo.cpmc.columbia.edu). The "PredictProtein" server includes: PHDsec (predicts secondary structure from multiple sequence alignments), PHDacc (predicts per residue solvent accessibility from multiple sequence alignments), PHDhtm (predicts the location and topology of transmembrane helices from multiple sequence alignments), GLOBE, TOPITS, MaxHom (dynamic multiple sequence alignment program which finds similar sequences in a database), EvalSec, COILS, ProSite (finds functional motifs in the sequence being investigated), SEG and ProDom (database of putative protein domains; searched with BLAST for domains corresponding to sequence being investigated) programs. In essence these servers allow the sequence of the ORP to be submitted for comparison to the sequences of proteins in the PredictProtein database. PredictProtein retrieves similar sequences and predicts secondary protein structure based on data for similar sequences. PredictProtein performs and displays the results of a "PROSITE" motif search, "ProDom" domain search, MAXHOM alignment header analysis, and provides information regarding

accuracy of the forgoing analyses. This prediction of secondary structure is performed by PredictProtein using a system of neural networks. The MAXHOM program produces a multiple sequence alignment file which serves as the input for the neural network system. The output of the MAXHOM analysis includes identification of aligned proteins, percentage of pairwise sequence identity, percentage of weighted similarity, number of residues aligned, number of insertions and deletions (indels), number of residues in all indels, length of aligned sequences and a short description of the aligned proteins. The preferred neural network for prediction of secondary structure is described in detail in: "Prediction of Protein Structure at Better than 70% accuracy" J. Mol. Biol., 1993, 232, 584-599, the entire disclosure of which is incorporated by reference. Prediction of solvent accessibility is also determined (PHDacc) in accordance with "The Analysis and Prediction of Solvent Accessibility in Protein Families" Proteins, 1994, 20, 216-226, the entire disclosure of which is incorporated by reference. The latter prediction provides values for the relative solvent accessibility. Prediction of helical transmembrane segments of the ORP is performed by the PHDhtm program. In this manner, the secondary structure (helix, sheet, loop) and location relative to the membrane (inside, outside, transmembrane) for the ORP under investigation is predicted with relative accuracy. Most preferably, the predicted topology for the transmembrane proteins is determined using PHDtopology and fold recognition is determined by predicted-based threading using PHDthreader. Again, the secondary structure predictive determinations are verified for accuracy using EvalSec. All of the computer programs used in the present invention can be accessed by the public and their disclosures are incorporated herein by reference. (see, embl-heidelberg.de/tmap_info.html).

Based on the results of the secondary structure prediction analysis, the sequences for the predicted seven transmembrane helices are determined. Next, the tertiary structure of the transmembrane helices are determined. Most preferably this is achieved in the present invention using the Swiss-Model 7TM Interface program and, preferably, BLAST (Basic local alignment search tool as described in J. Mol. Biol. 215:403-410, the entire disclosure of which is incorporated herein by reference). To begin, the complete sequence of the ORP under investigation is input in the BLAST program which then determines the most appropriate modeling template to be used in the tertiary structure investigation. The

modeling template will be that protein (of known primary, secondary and tertiary structures) having the highest primary sequence homology with the ORP to be investigated. In other words, using BLAST the primary sequence of the ORP under investigation is compared to the sequences of proteins in the 7TM subset of the SWISS-PROT database.

After the modeling template has been selected, the sequences of the helical regions are displayed and the sequences of the helices of the ORP under investigation (as determined in the secondary structure analysis step of the present invention) are input (Swiss-Model 7TM Interface program). That is, the helical regions of the template are aligned with the helical regions of the ORP under investigation. The comparison yields a prediction of the tertiary structure (3D in space) of the ORP being investigated on an atom-by-atom basis. The preferred protocol for this step takes into consideration energy minimization and the like as described in: "Promod and Swiss-Model: Internet-based Tools for Automated Comparative Protein Modeling, Biochem. Soc. Trans. V. 24 274 1996; Large-Scale Comparative Protein Modeling, Proteome Research: New Frontiers in Functional Genomics 177 1997; Swiss-Model and the Swiss-PDBviewer: an Environment for Comparative Protein Modeling, Electophoresis, V. 18 2714 1997; Automated Modeling of the Transmembrane Region of G-Protein Coupled Receptor by Swiss-Model, Receptors and Channels v. 4 161 1996; Protein Modeling by email, Bio/Technology V. 13 658 1995, the disclosures of which are incorporated by reference. (The preferred modeling software programs which can be used in the present invention have a high degree of sophistication. For example, ProMod is a knowledge-based approach to predictive structure determination. It requires at least one known 3D structure of a related protein and good quality sequence alignments; the degree of sequence identity affects the accuracy of the predictive structure. In ProMod, there is a superposition of related 3D structures. A multiple alignment with the sequence under investigation is made. A framework for the new sequence is made and any missing loops are rebuilt. The backbone of the structure is completed and corrected if required. Side chains are corrected and rebuilt. The resultant structure is verified and packing is checked. The structure is then refined by energy minimization and molecular dynamics considerations.)

The tertiary structures of the helices of the ORP under investigation are thus

determined and may be viewed stereoscopically using a program such as Insight II or Swiss PDB-viewer or the like. Next, a ligand, preferably one which is known to bind to the ORP under investigation, is selected. A number of assays may be used to determine high general binding affinities of various ligands for the ORP under investigation. The molecular structure of the ligand is then input to the Insight II program, i.e. the "tertiary or 3D structures of ORP helices and the ligand are input. Next, the most probably geometrical binding domains of the ORP under investigation and the ligand are determined, preferably using the Global Range Molecular Modeling program (GRAMM) which utilizes geometric recognition algorithms. As will be understood by those skilled in the art, GRAMM is a program for protein docking; no specific information about the binding sites is required. It performs a six-dimensional search through the relative translations and rotations of molecules. It takes an empirical approach to smoothing the intermolecular energy function by changing the range of atom-atom potentials. It allows the user to locate the area of the global minimum of intermolecular energy for structures of different accuracy. Insight II may then be used to calculate the energy distribution and reaction forces between the ligand and the geometrically predicted domains. The most probably overall binding domains are thus determined.

Polypeptides are then synthesized corresponding to these binding domains using conventional synthesis technologies. The polypeptides are then applied to the surface of a transducer, preferably one fabricated using thin film (semiconductor) techniques, as will be known to those skilled in the art. Briefly, with reference to Figure 3, biosensor 10 is seen having transducers 12 coated with polypeptide layer 14. Transducer 12 is preferably a piezoelectric quartz crystal-based device. A mass change will occur if a ligand binds to the polypeptide layer resulting any a measurable frequency change in the quartz crystal frequency, allowing detection of ligand binding. The success and efficience of the transducer can be determined, including by comparing the sensor's response to the ligand and other molecules.

Examples:

The following examples are intended to further illustrate the present invention.

A G-Protein Coupled Receptor database was accessed and the sequence of an ORP of known primary sequence, but unknown secondary and tertiary structures was retrieved (SWISS-PROT:P30955) as shown in Figure 4. It consists of 330 amino acids and has a molecular weight of 35197 daltons. The secondary structure was predicted and its accuracy verified through the use of MaxHom, PHDsec, PHDacc, PHDhtm, PHDtopology, PHDthreader and EvalSec. The transmembrane sequence regions were thus obtained.. A BLAST assisted template was then selected: Neuropeptide Y1 receptor (Homo sapiens). Trimethylamine was selected as the ligand. Using GRAMM, several possible binding domains were identified and corresponding polypeptides were generated. In Figure 5, (poly)peptide B1 designed in accordance with the present invention illustrates better response for trimethylamine than another (poly)peptide Pb2.